# Large Data Visualization

Kwan-Liu Ma
Department of Computer Science
University Of California at Davis



NASA 04/21/2010

# Outline

- VIDi Research Group
- Visualization highlights

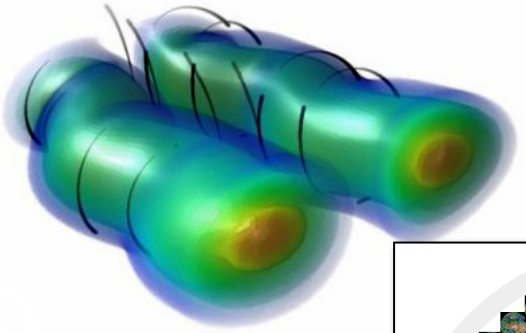
---

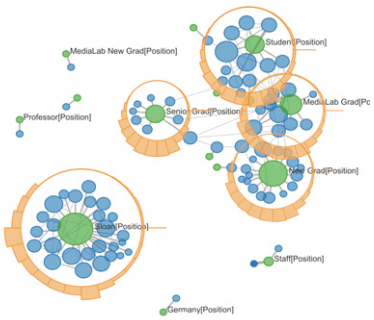
- Large data visualization
- Institute for Ultra-Scale Visualization
- Visualization highlights
  - Knowledge assisted data reduction & vis
  - In situ visualization
  - Explorable images

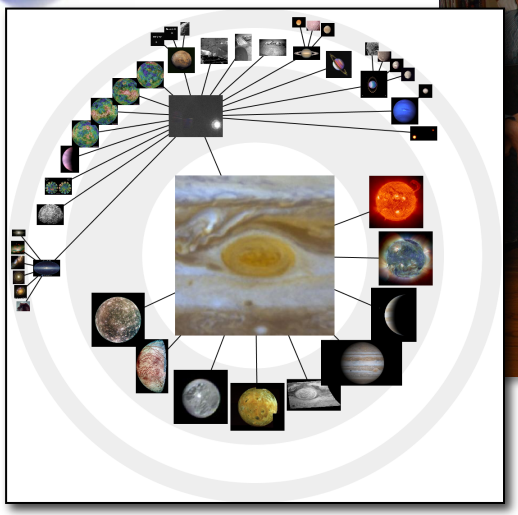
---

# ViDi

Visualization and Interface Design Innovation

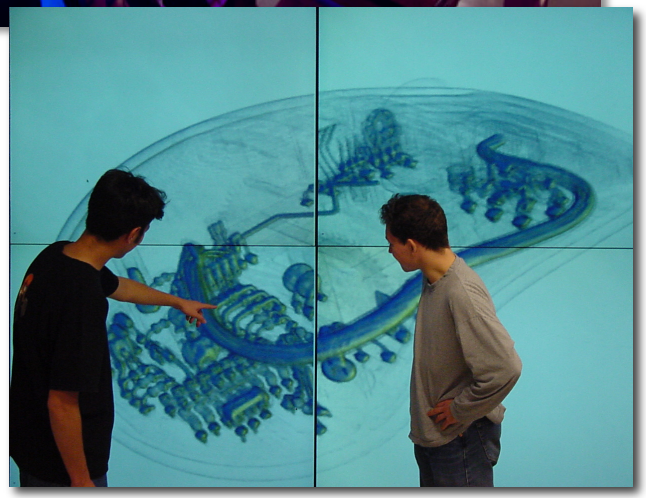




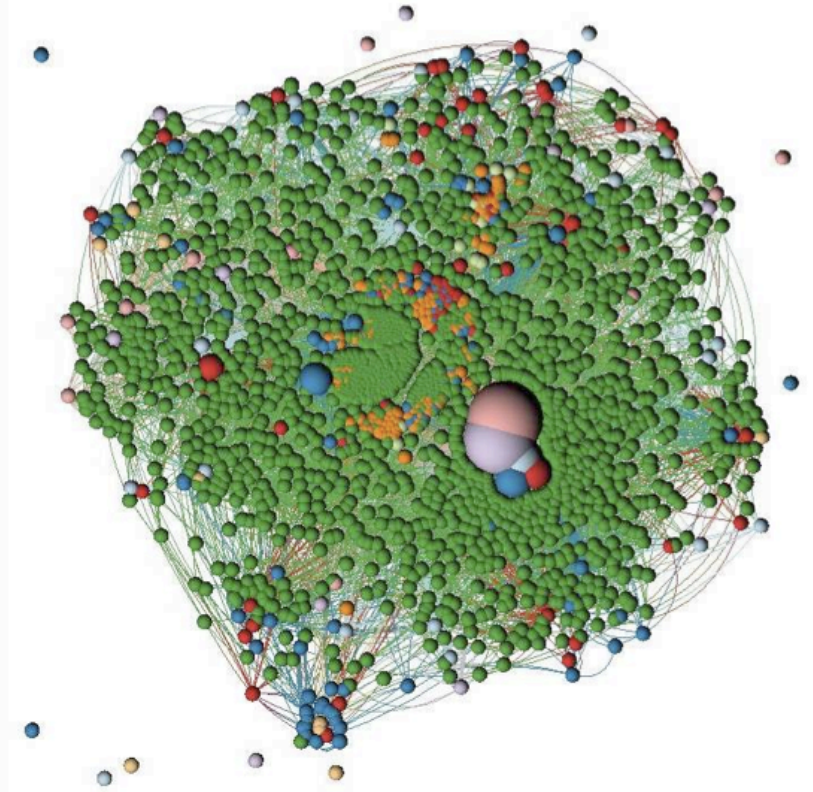



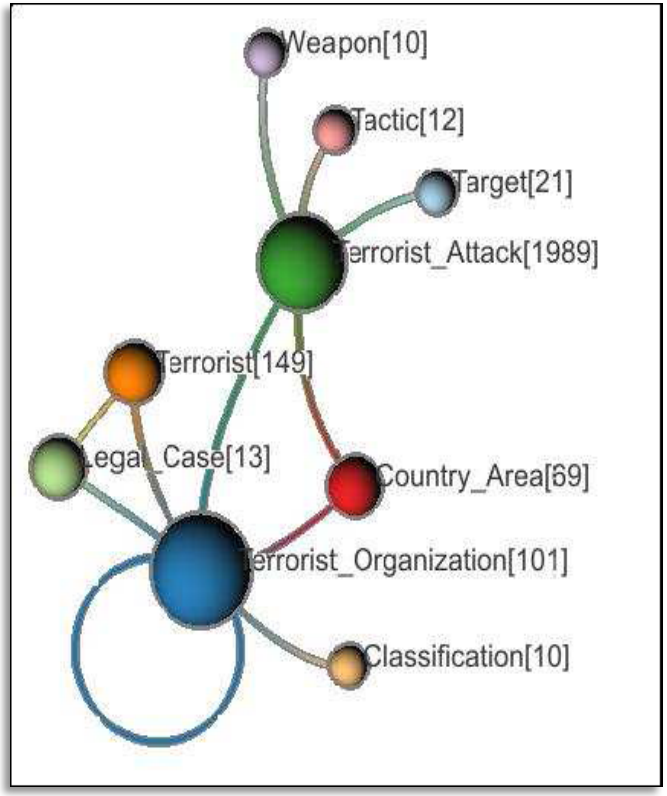
# Research Projects

- Flow visualization
- Volume data visualization
- Large data visualization
- Parallel visualization
- Biomedical data visualization
- Graph and network visualization
- Visual analytics
- Software visualization
- Performance visualization
- Intelligent visualization
- Remote, collaborative visualization
- Advanced interfaces and interaction techniques



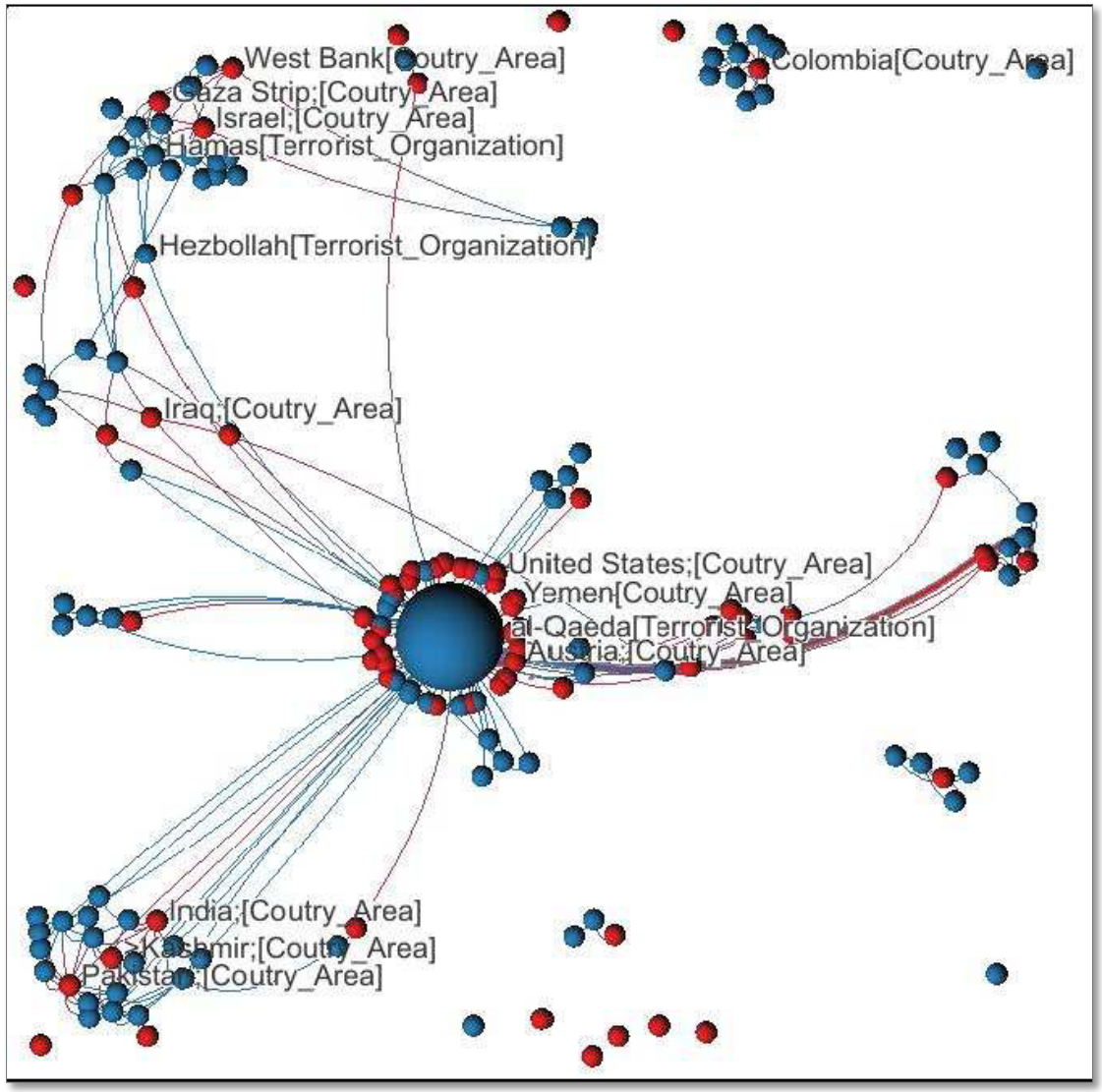
# Terrorist Network Analysis





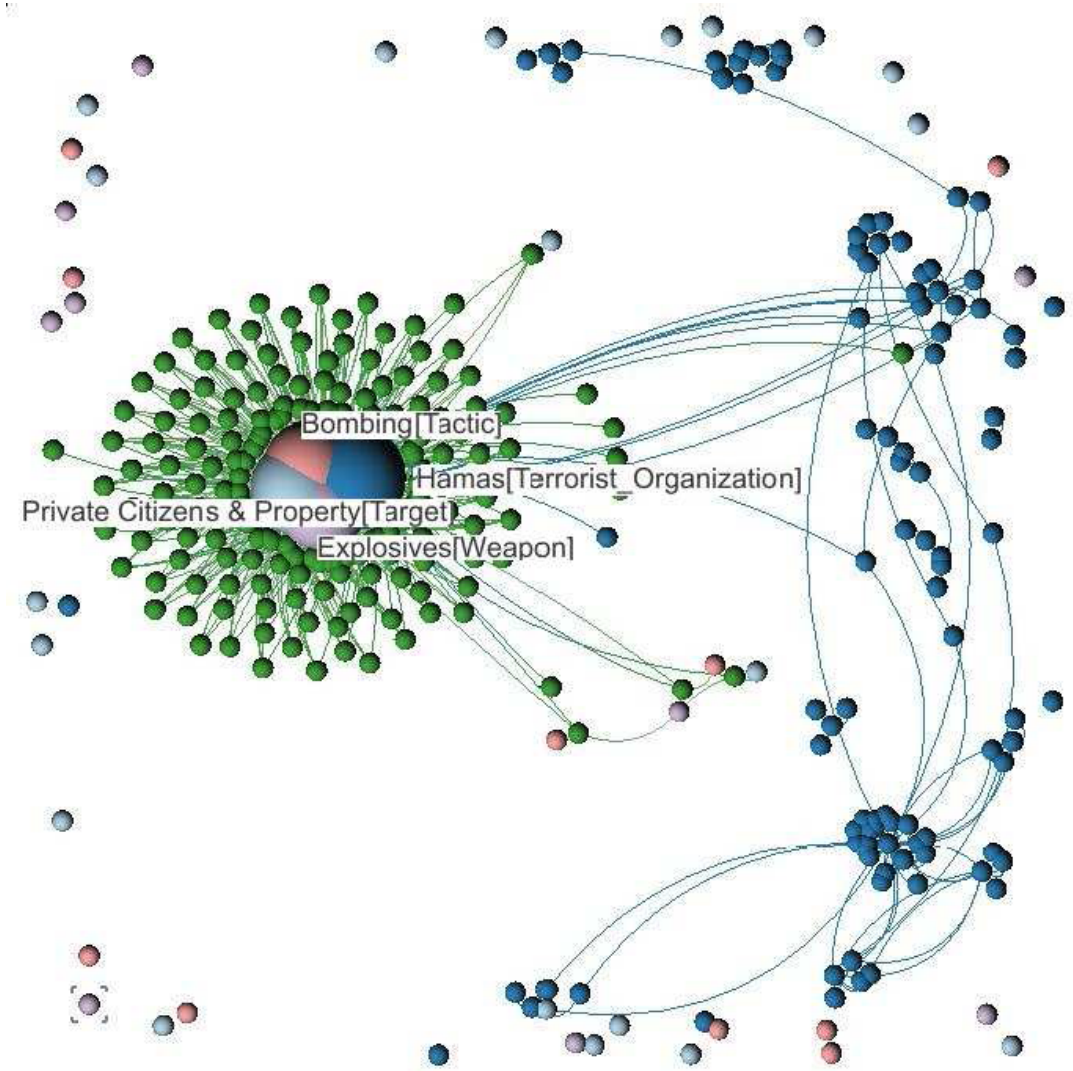

Tasks: Identify the key terrorist organizations.
Find the relationships among the organizations
Characterize their behaviors

# Terrorist Network Analysis



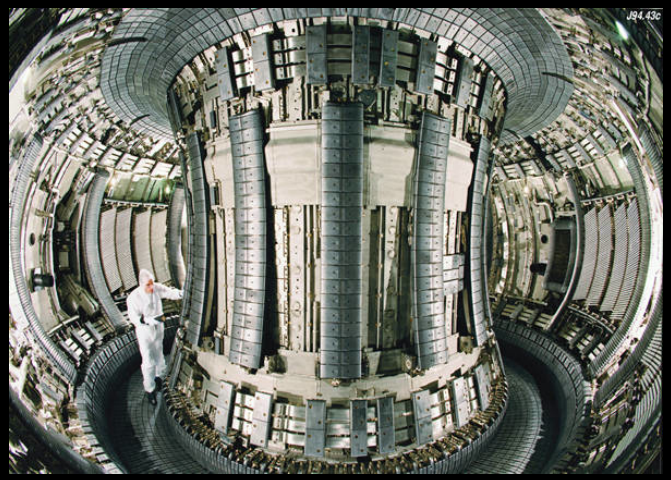

Organization clusters are developed based on locations

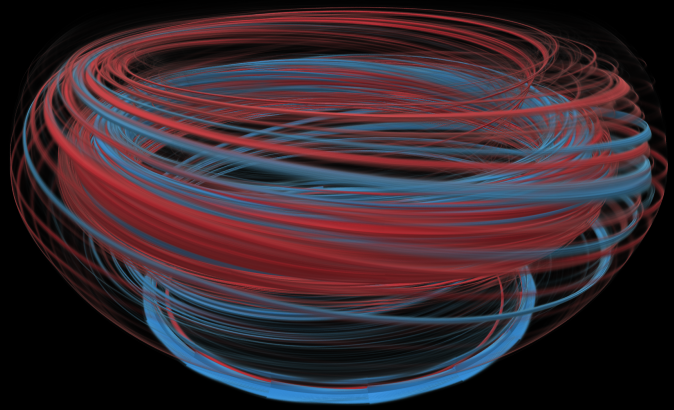
# Attack Characteristics

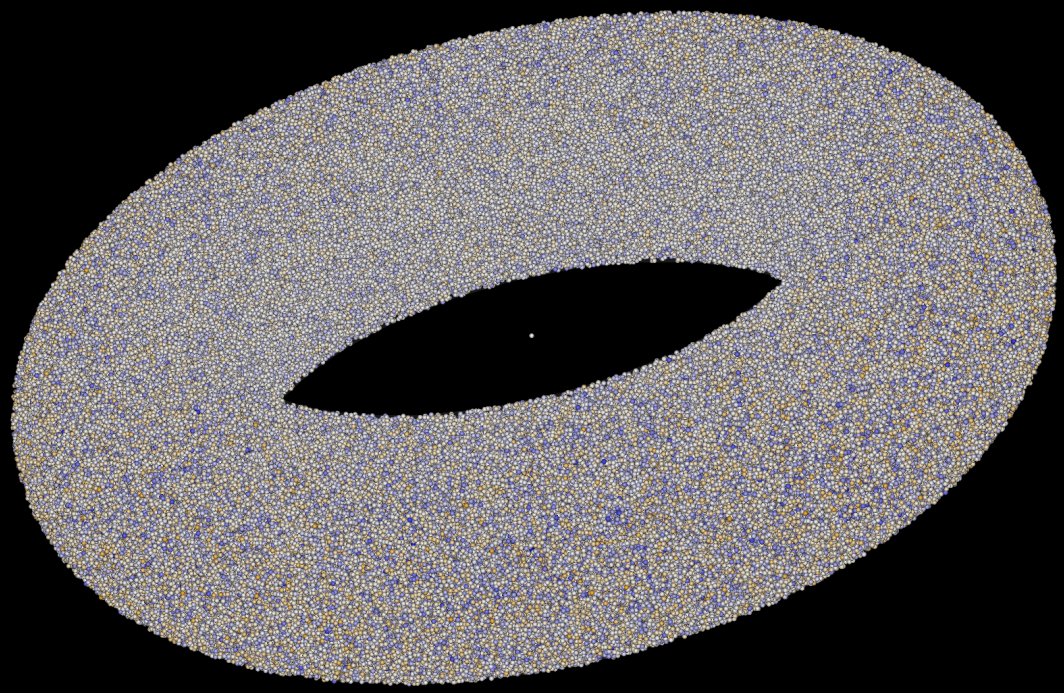


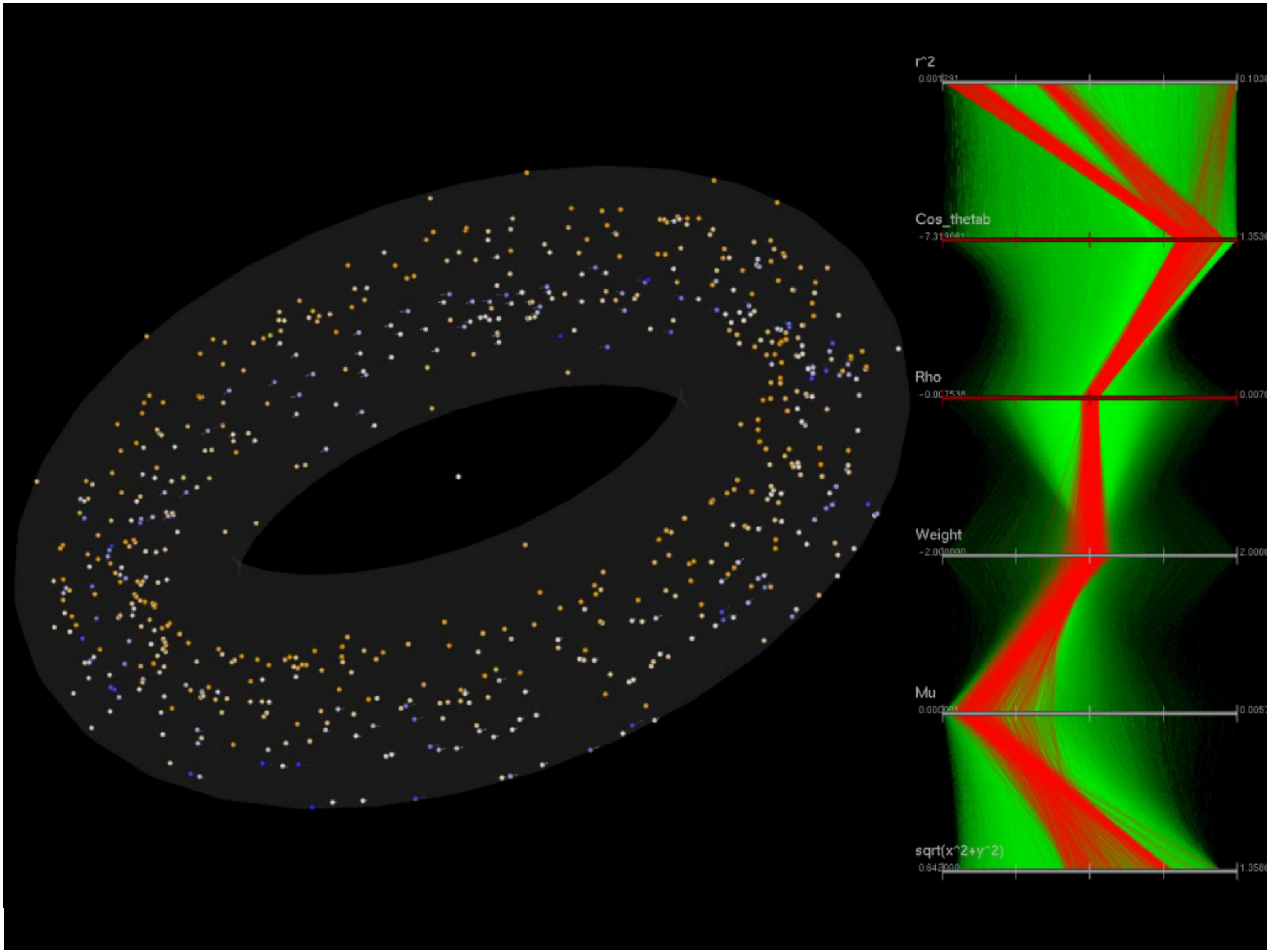

Add attack target, tactic and weapon. Add Hamas and its attacks in 2005.

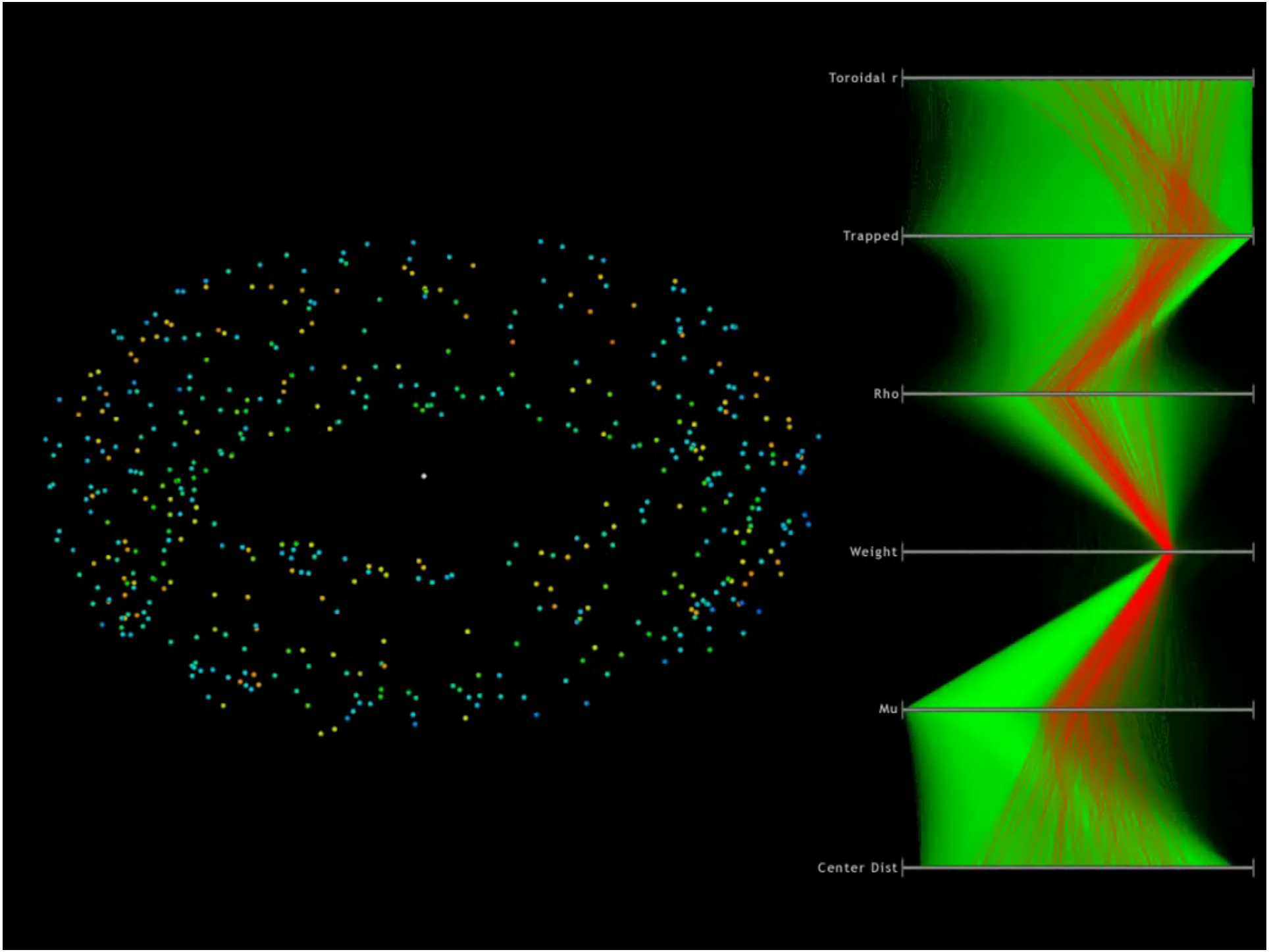
# Multidimensional Particle Data Visualization





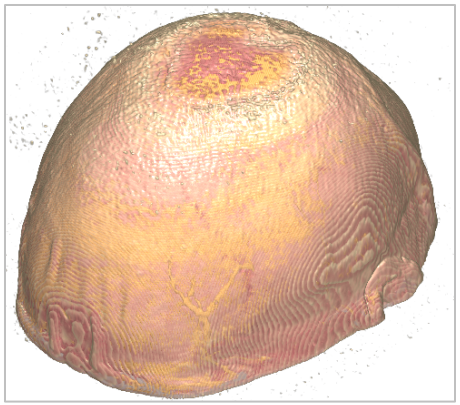


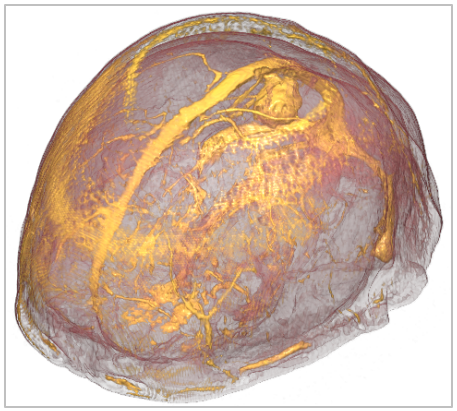

r^2
Cos_thetab
Rho
Weight
Mu
sqrt(x^2+y^2)

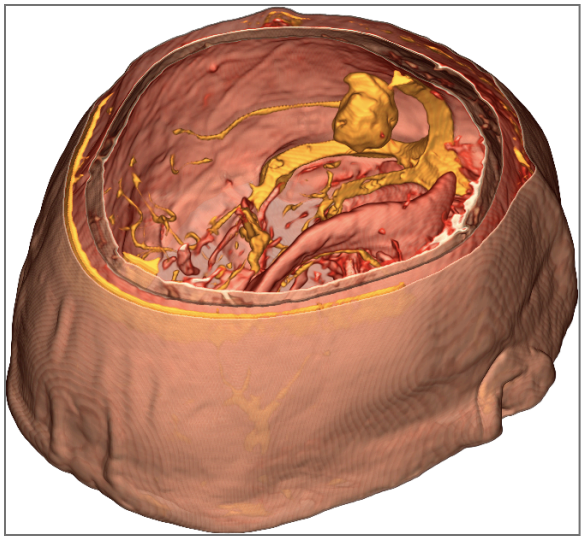

Toroidal r
Trapped
Rho
Weight
Mu
Center Dist

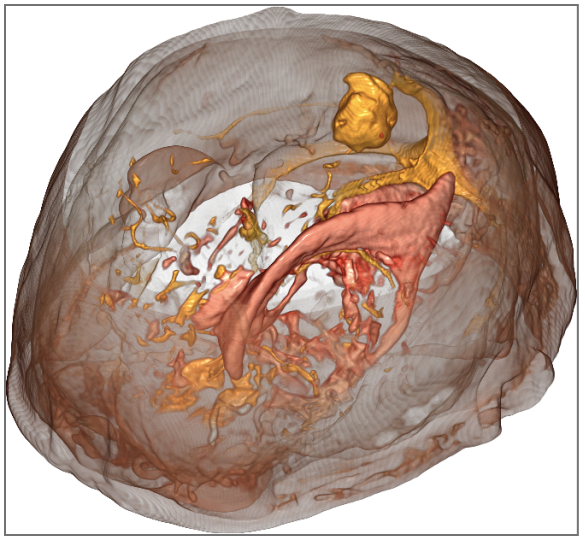
# Biomedical Data Visualization

- Novel volume data classification techniques
- Multimodal volume data visualization
- Biological network visualization

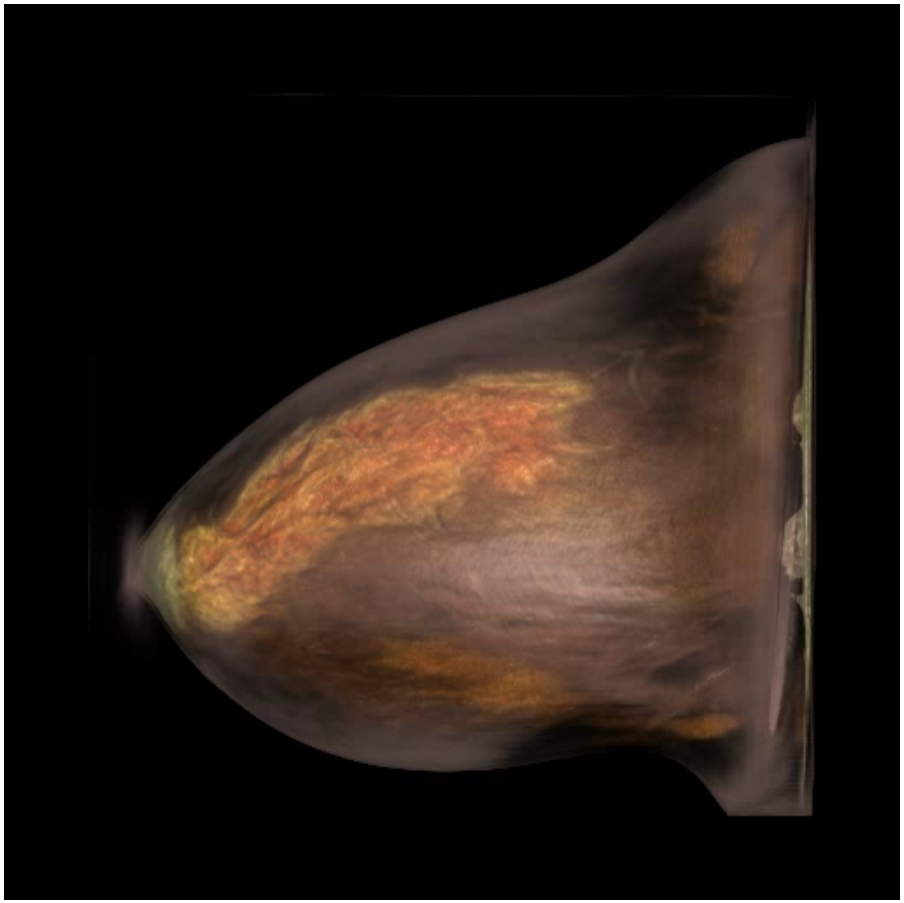


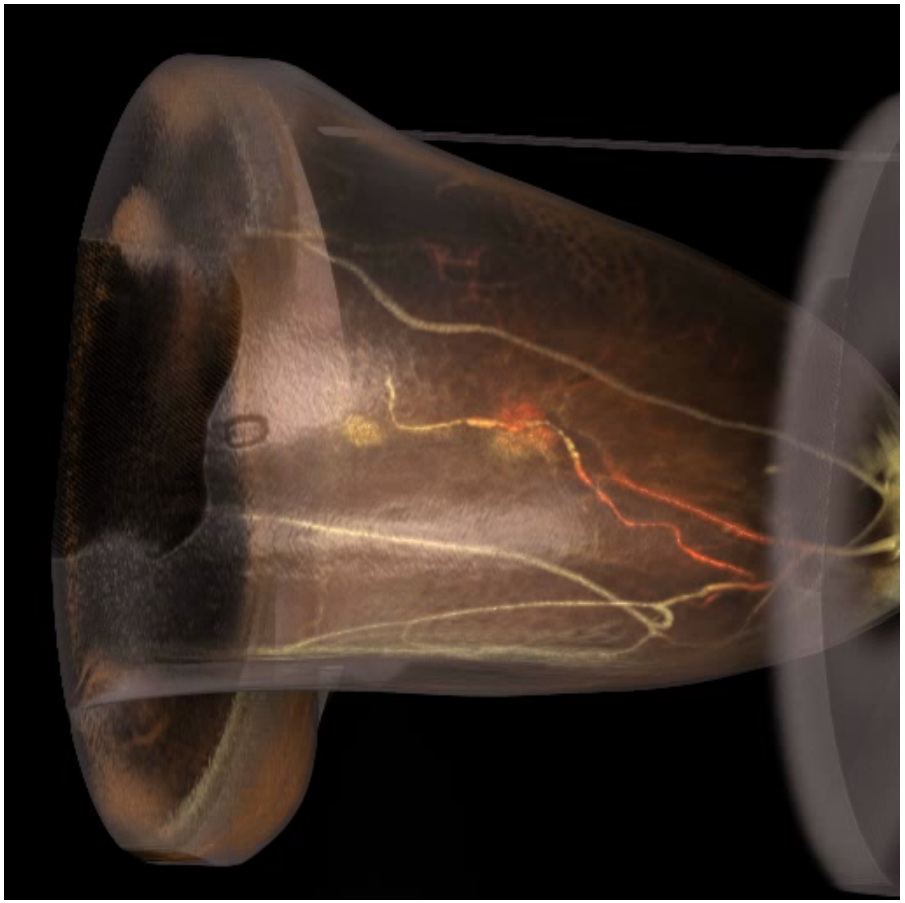




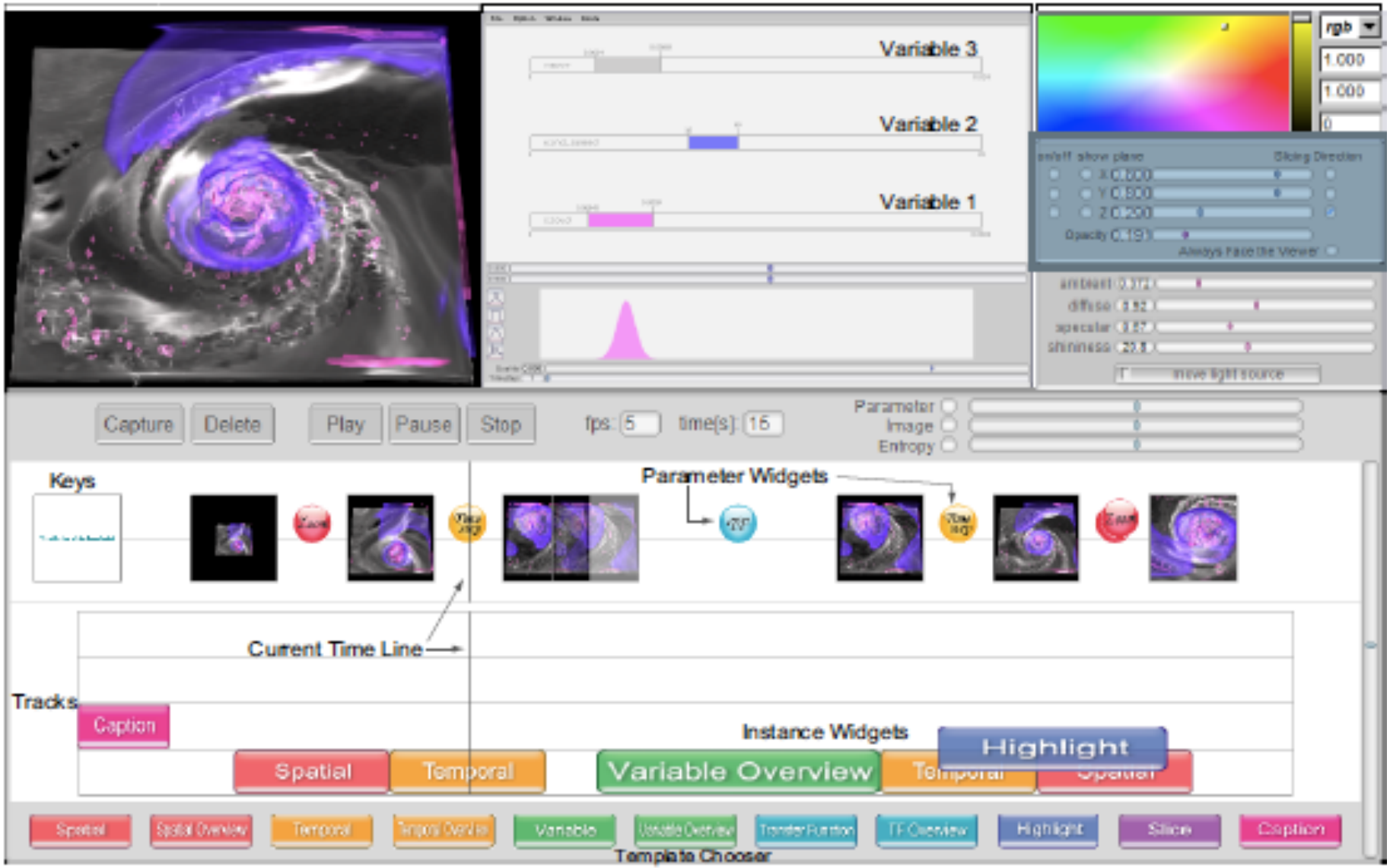


# Occlusion Based Visualization





# AniViz


Variable 3
Variable 2
Variable 1
rgb
Capture
Delete
Play
Pause
Stop
fps: 5
time[s]: 15
Parameter
Image
Entropy
Keys
Parameter Widgets
Current Time Line
Tracks
Caption
Instance Widgets
Highlight
Spatial
Temporal
Variable Overview
Template Chooser

# Large Data Visualization

State of the art computing technologies:

- Simulations
- Supercomputing facilities
- Visualization solutions

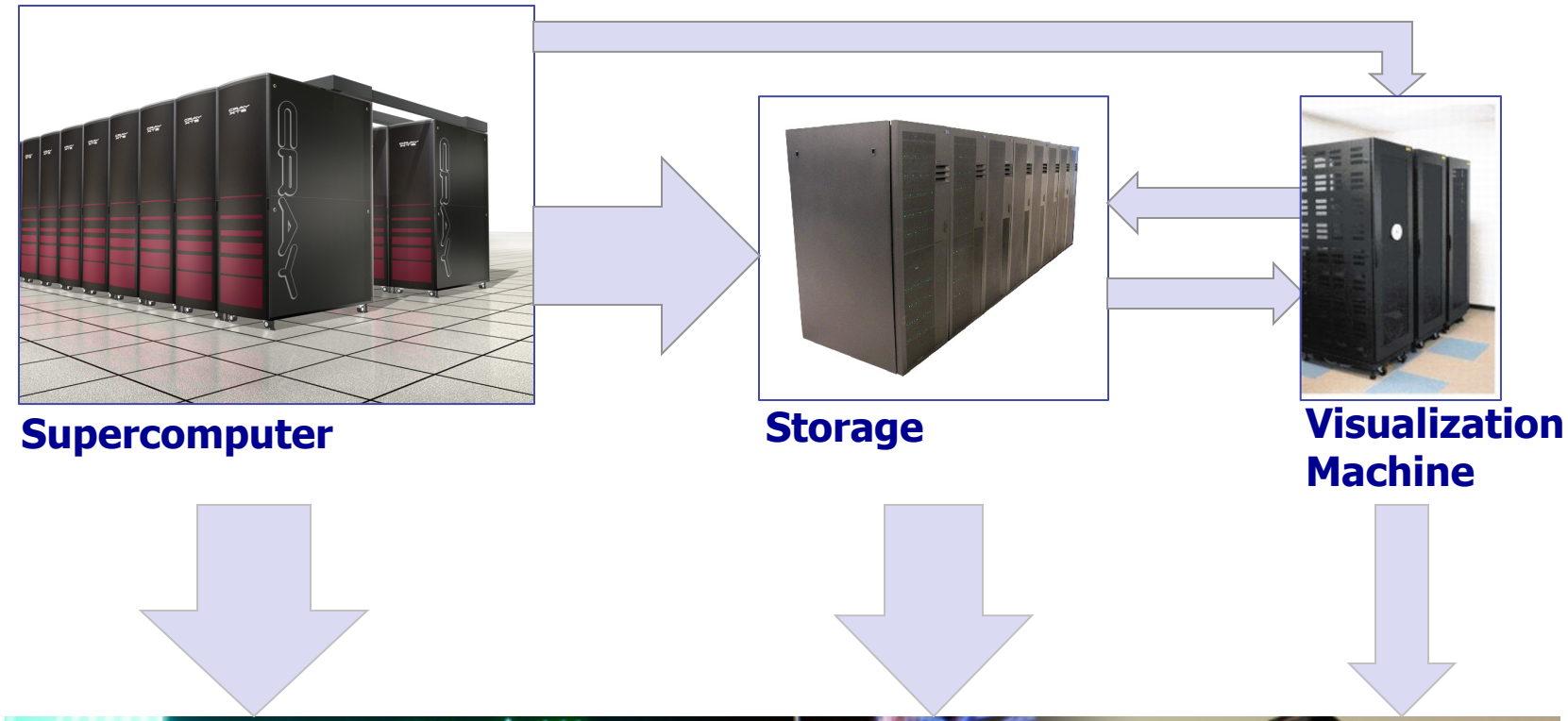
# Top Supercomputers

| Rank | Site | Computer/Year Vendor | Cores | $R_{max}$ | $R_{peak}$ | Power |
|---|---|---|---|---|---|---|
| 1 | Oak Ridge National Laboratory<br>United States | Jaguar - Cray XT5-HE Opteron Six Core 2.6 GHz / 2009<br>Cray Inc. | 224162 | 1759.00 | 2331.00 | 6950.60 |
| 2 | DOE/NNSA/LANL<br>United States | Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband / 2009<br>IBM | 122400 | 1042.00 | 1375.78 | 2345.50 |
| 3 | National Institute for Computational Sciences/University of Tennessee<br>United States | Kraken XT5 - Cray XT5-HE Opteron Six Core 2.6 GHz / 2009<br>Cray Inc. | 98928 | 831.70 | 1028.85 | |
| 4 | Forschungszentrum Juelich (FZJ)<br>Germany | JUGENE - Blue Gene/P Solution / 2009<br>IBM | 294912 | 825.50 | 1002.70 | 2268.00 |
| 5 | National SuperComputer Center in Tianjin/NUDT<br>China | Tianhe-1 - NUDT TH-1 Cluster, Xeon E5540/E5450, ATI Radeon HD 4870 2, Infiniband / 2009<br>NUDT | 71680 | 563.10 | 1206.19 | |
| 6 | NASA/Ames Research Center/NAS<br>United States | Pleiades - SGI Altix ICE 8200EX, Xeon QC 3.0 GHz/Nehalem EP 2.93 Ghz / 2009<br>SGI | 56320 | 544.30 | 673.26 | 2348.00 |

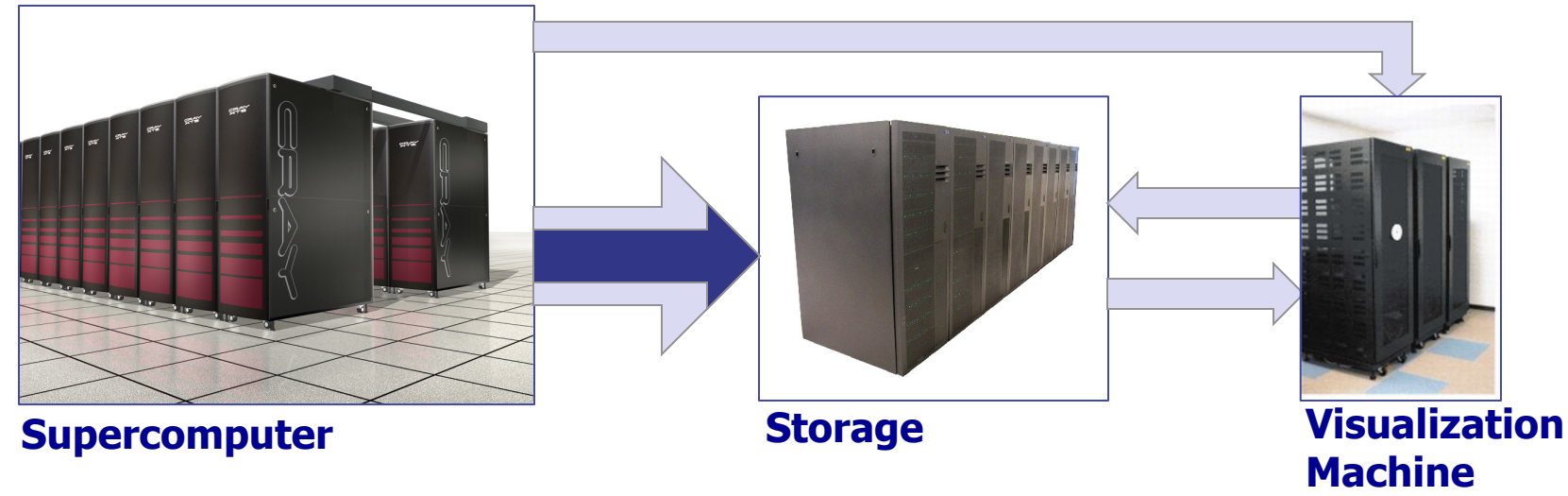
Turbulent nuclear simulations: 300TB
Climate simulations: 345TB
Turbulent combustion simulations: 250TB
Parkinson's disease: 50TB
Gating mechanisms of membrane proteins: 10TB

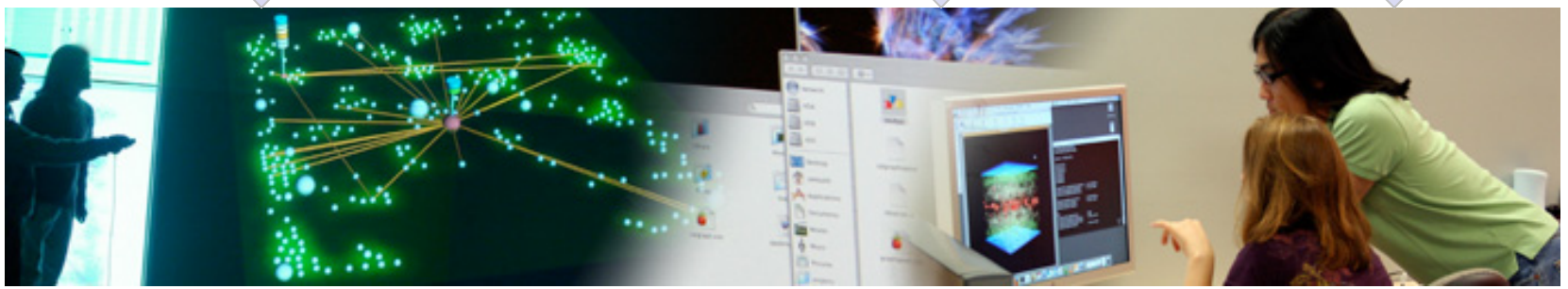
# Supercomputer Performance

# Large Data Visualization



# Large Data Visualization



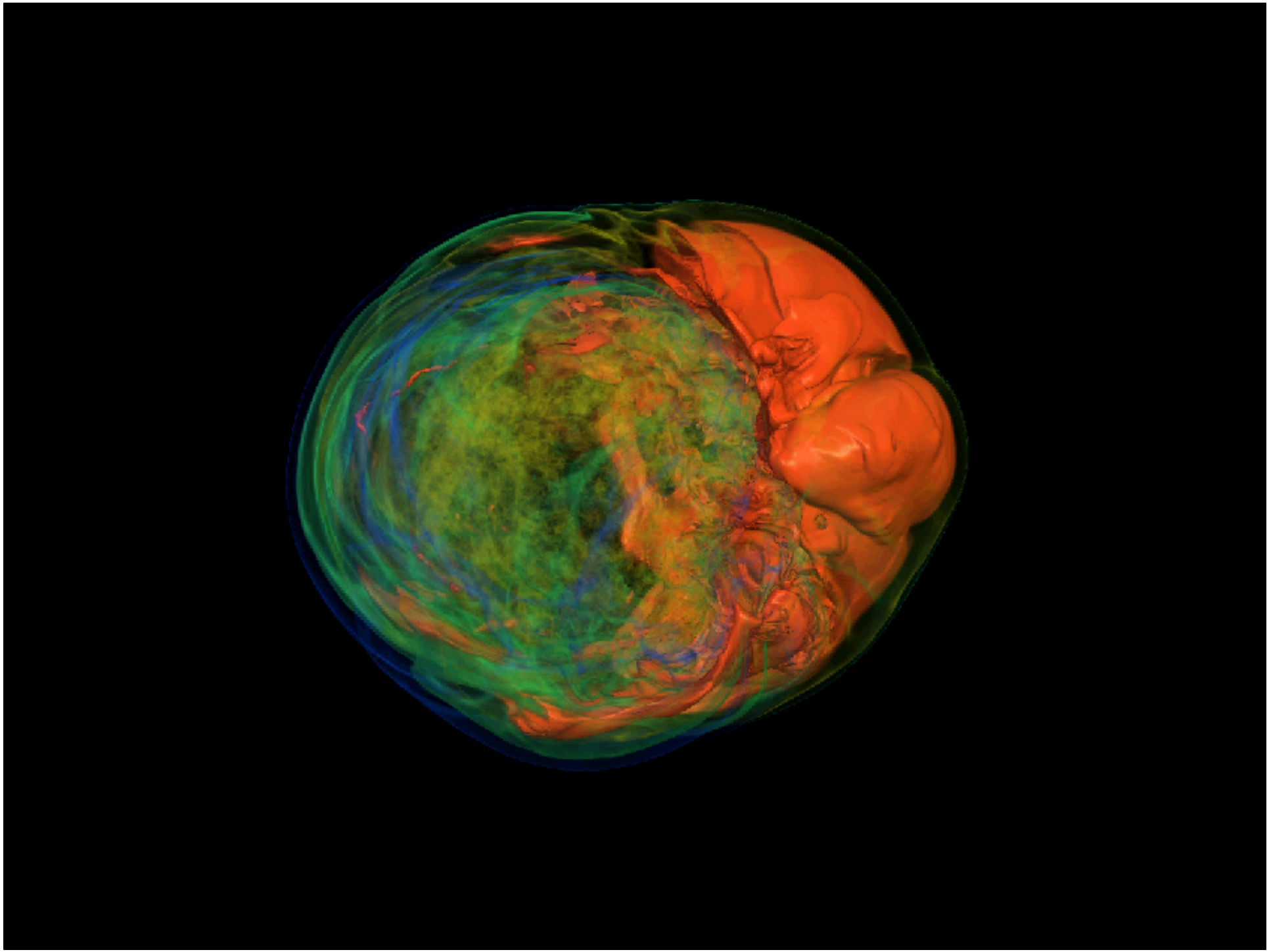



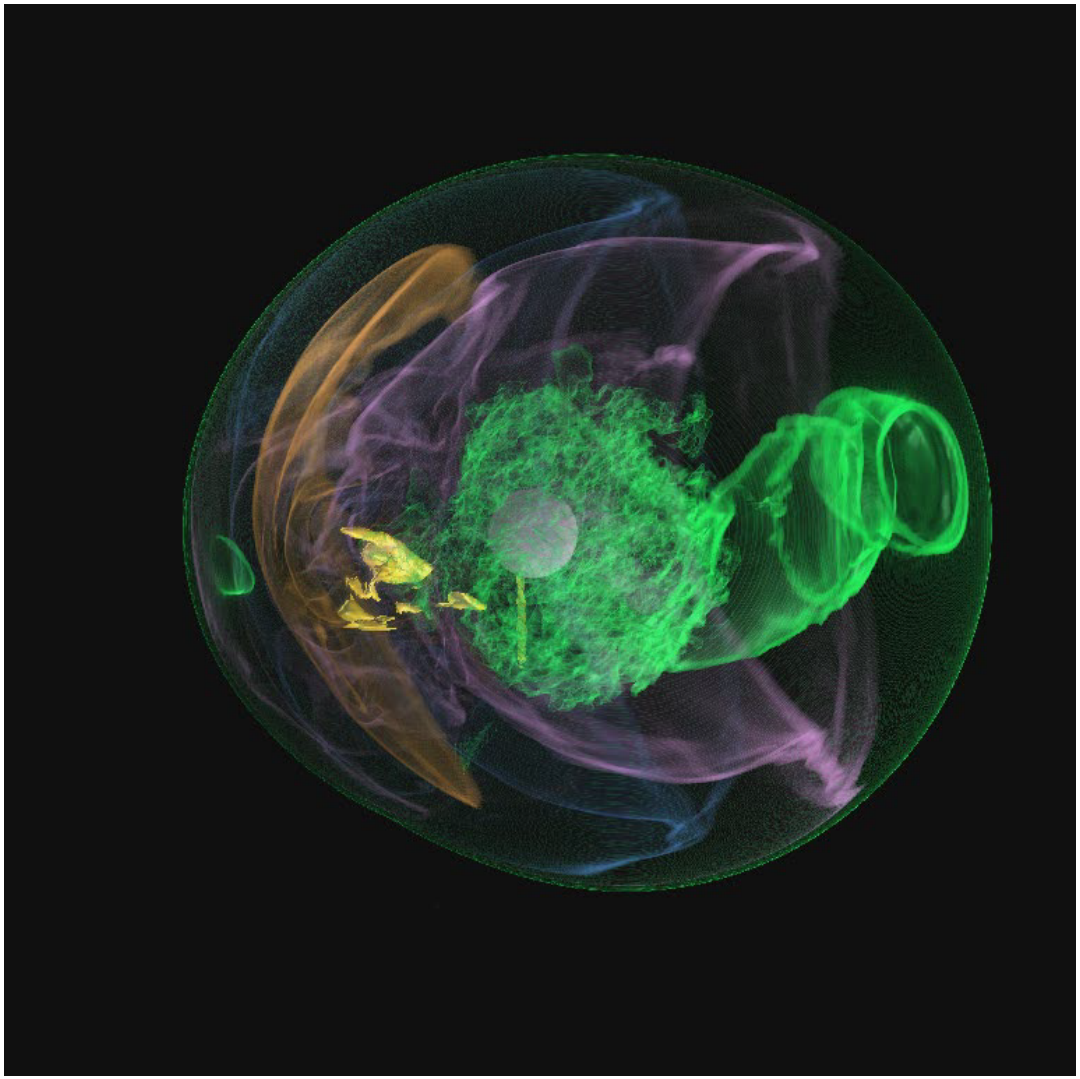
# Large Data Visualization

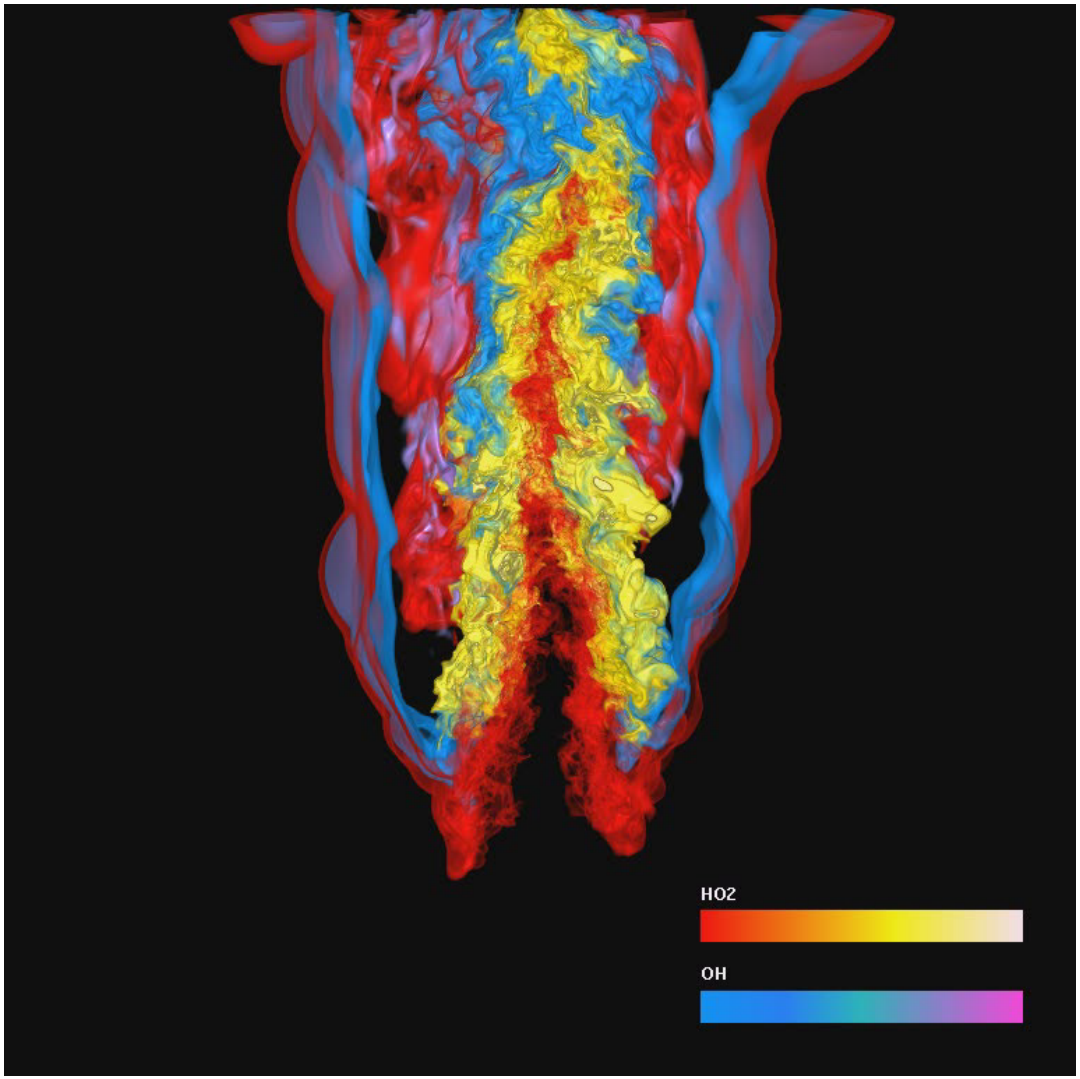
- Data reduction
  - Subset/skipping
  - Compression
  - Physically based/**knowledge assisted** feature extraction and **data reduction**
- Parallel visualization
  - Post-processing
  - Co-processing
  - **In situ**
- Remote visualization
  - Transfer data
  - Transfer extracts/geometry
  - Transfer **images**


- Five-year project sponsored by U.S. DOE SciDAC involving investigators from 4 universities and 2 national labs
- **Advance the visualization technology to enable knowledge discovery at extreme scale**, foster awareness of and communication about new visualization technologies, and put these technologies into the hands of application scientists
- Science application driven projects
- Enable scientists to see the previously unseen and more effectively communicate with others their findings
- http://ultravis.org

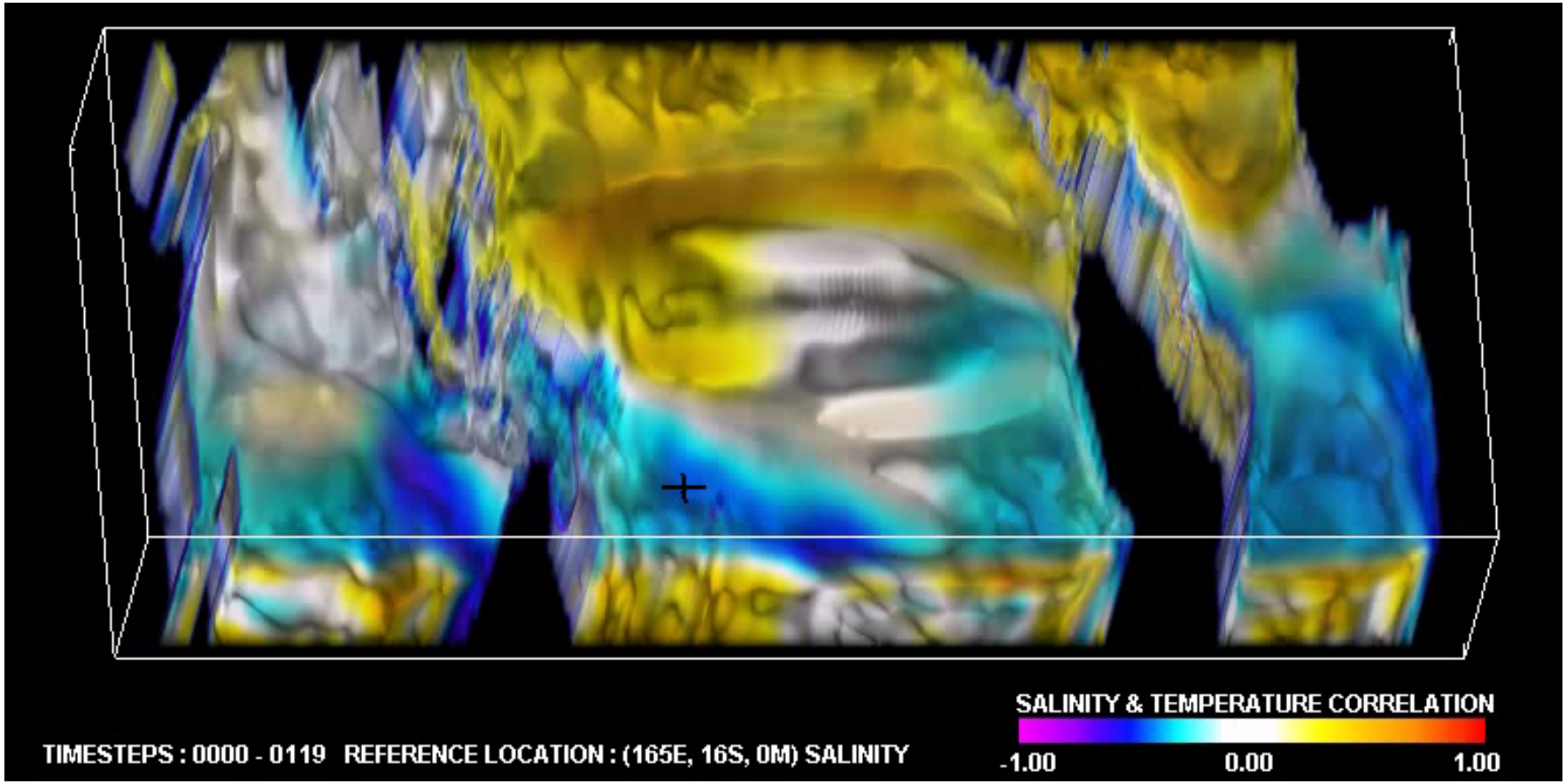

- Established collaborations with scientists from various application areas
- Created and demonstrated new concepts and technologies for visualization-driven knowledge discovery
- From 2006-2009, published over 100 research papers, organized over twenty workshops/tutorials/panels, gave over 100 invited talks including several keynote speeches, graduated 14 PhD students, and received over ten awards from investigators' respective societies
- Delivered open-source of toolkits and libraries for high-performance, high quality visualization and graphics
- Drew Wired, MIT Technical Reviews, NY Times, HPCWire, Discovery, and technical journals to feature our work



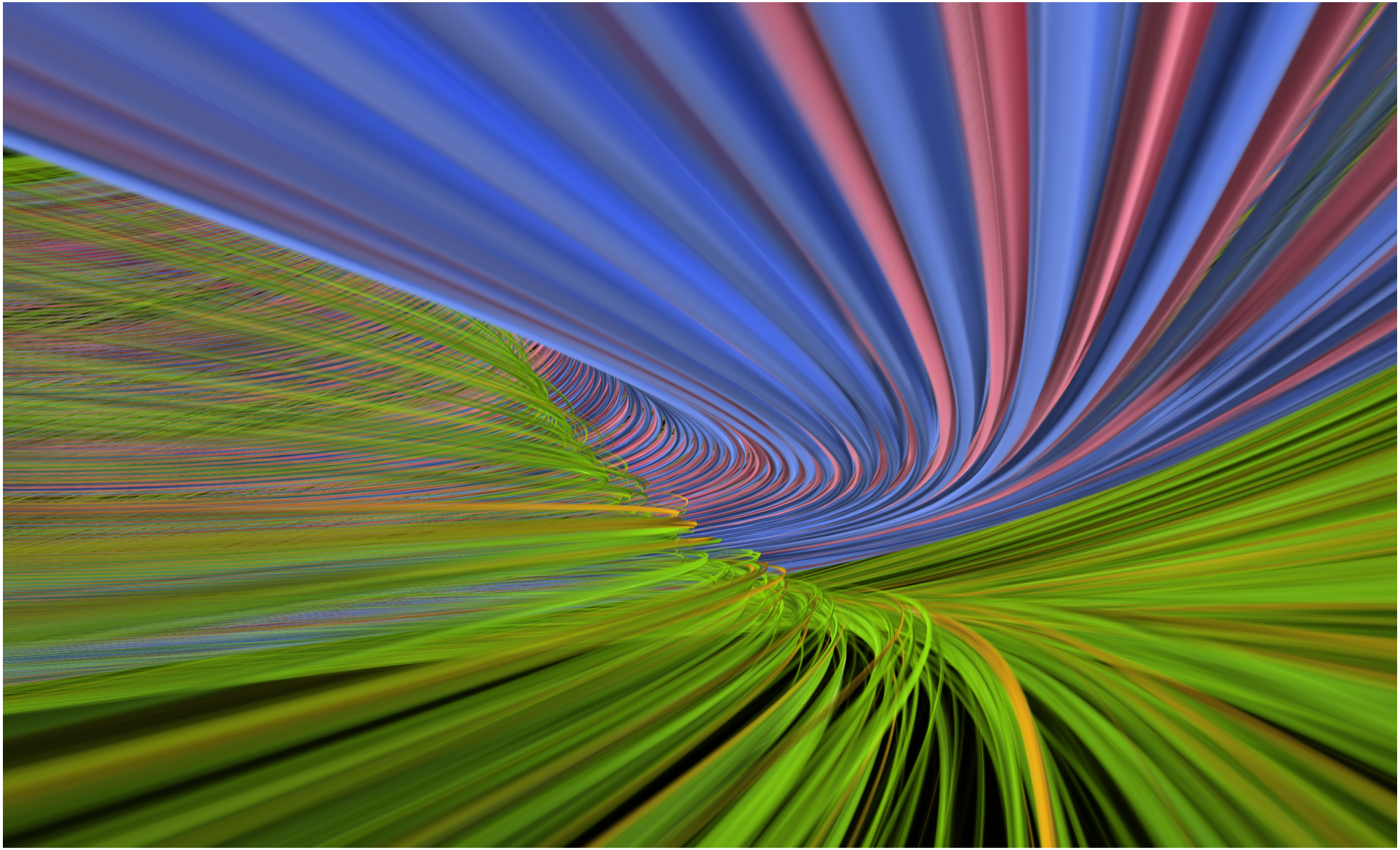


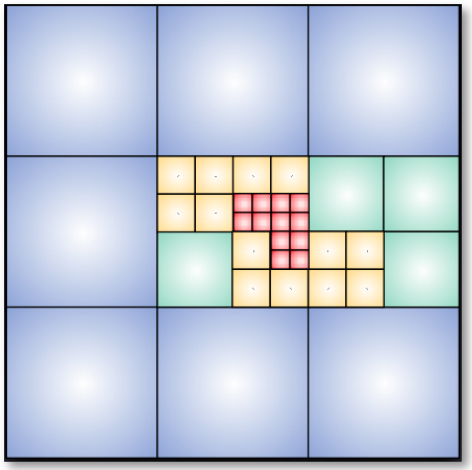

HO2
OH

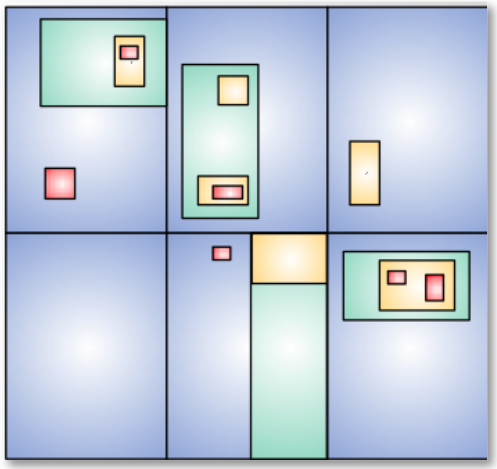
# 4D Correlation Analysis

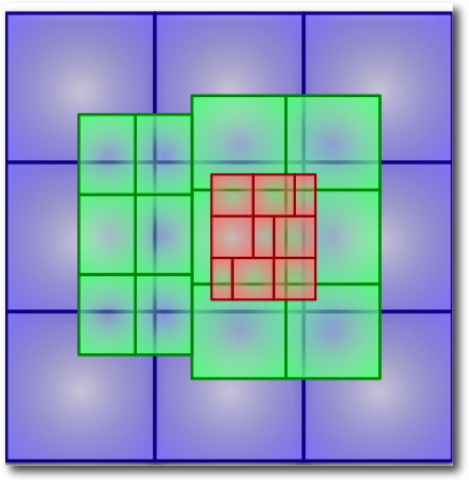


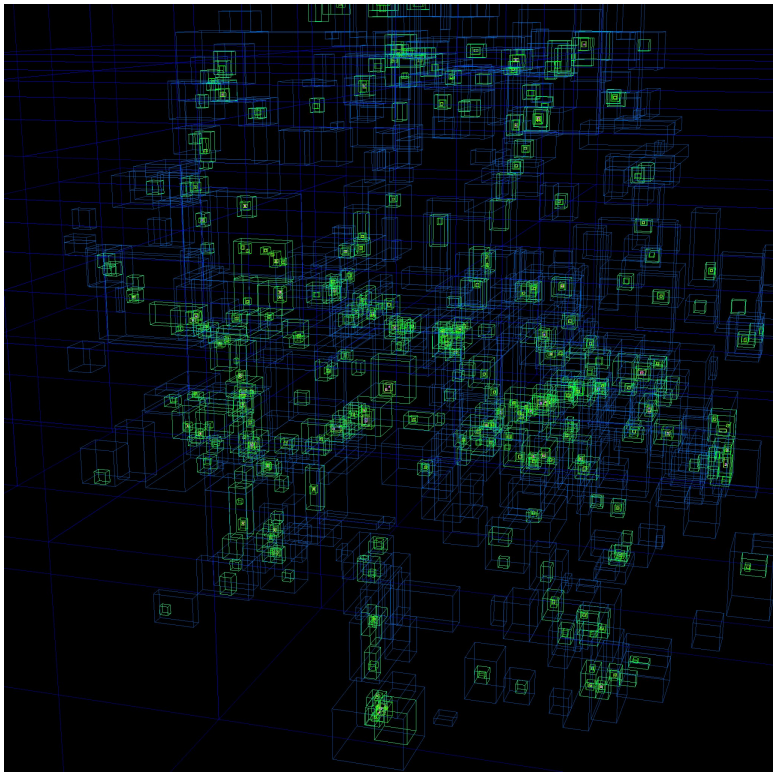

- On the fly calculation of correlation volume
- Interactive time-dependent 3D correlation analysis

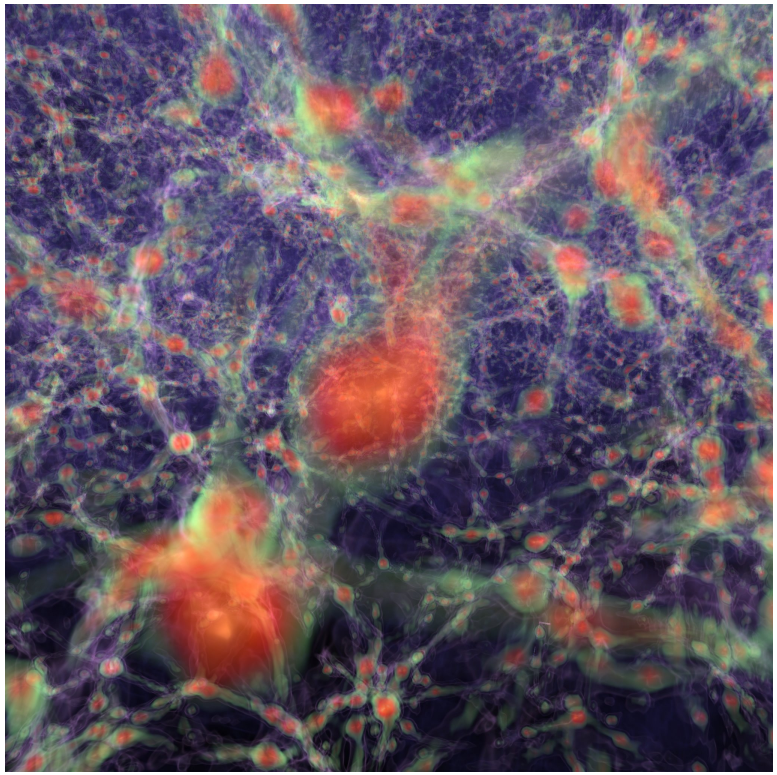






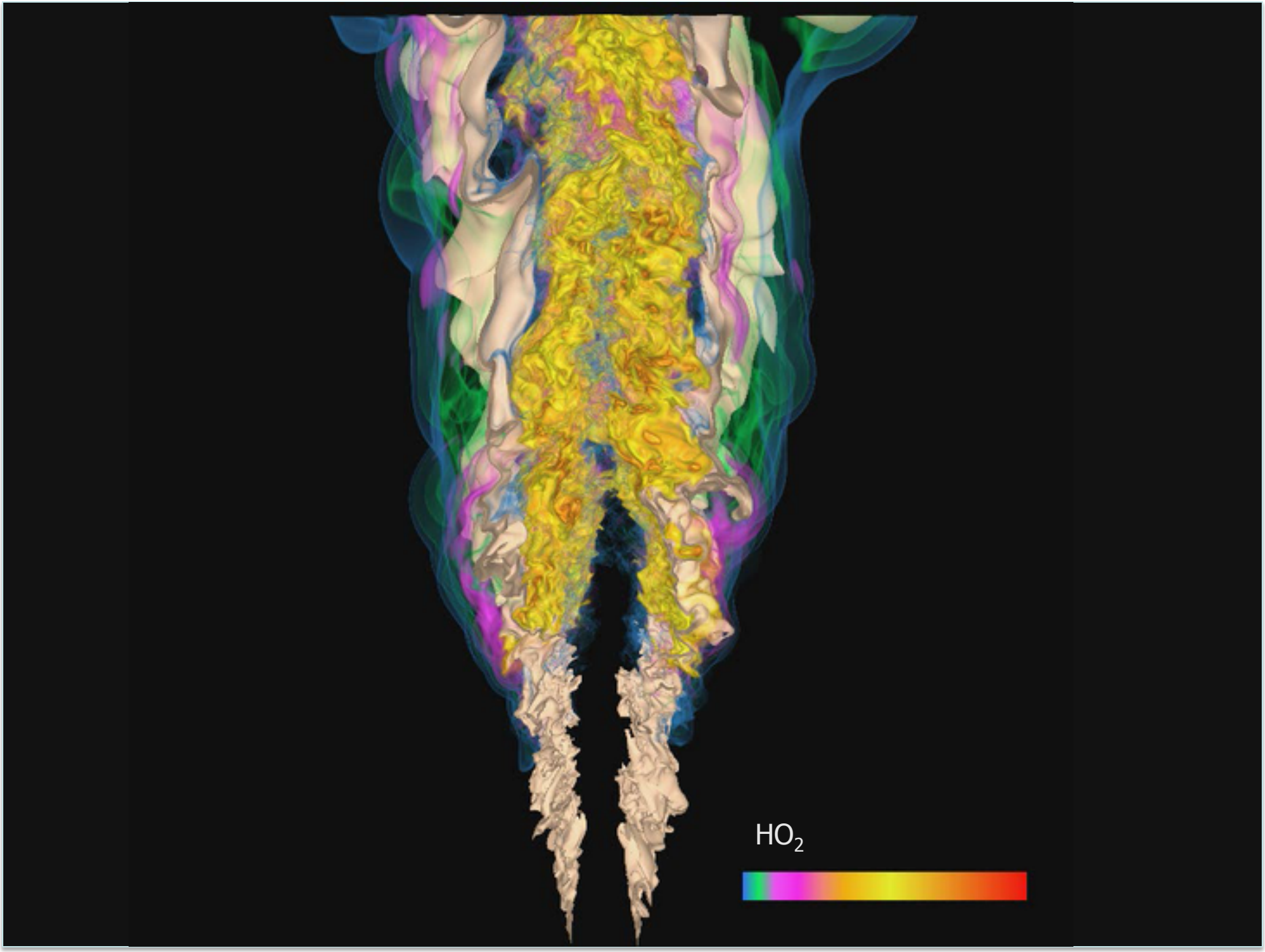




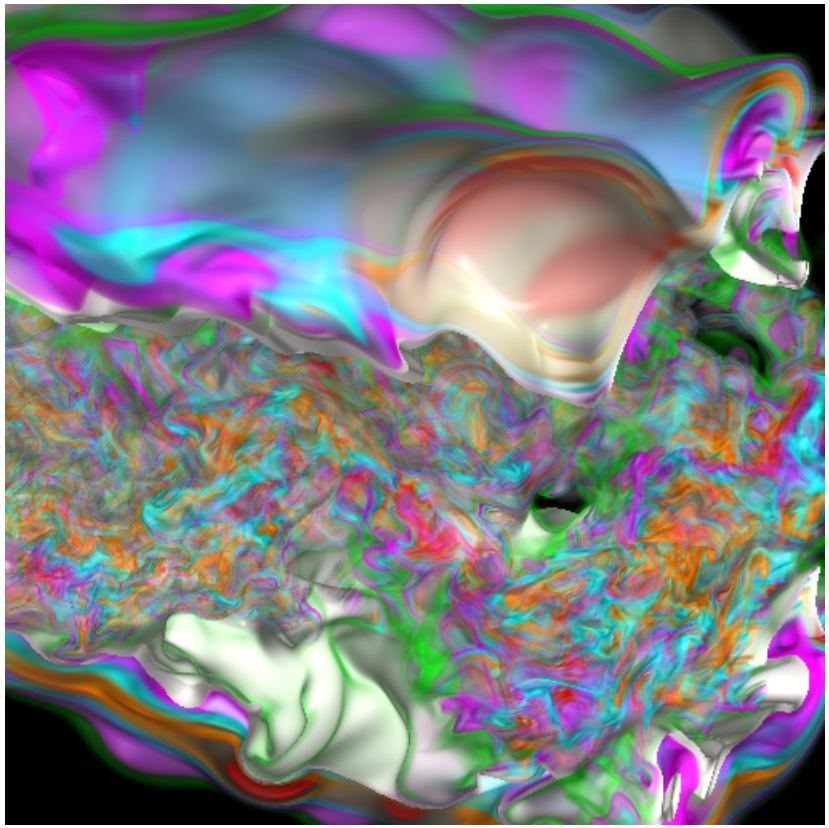


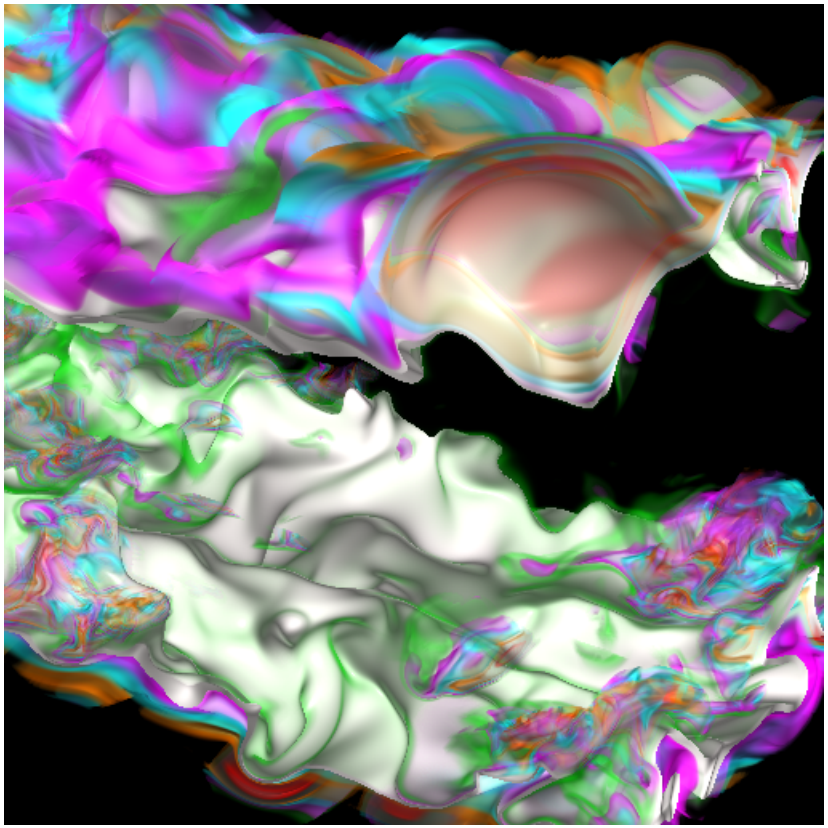
# Knowledge-Assisted Visualization

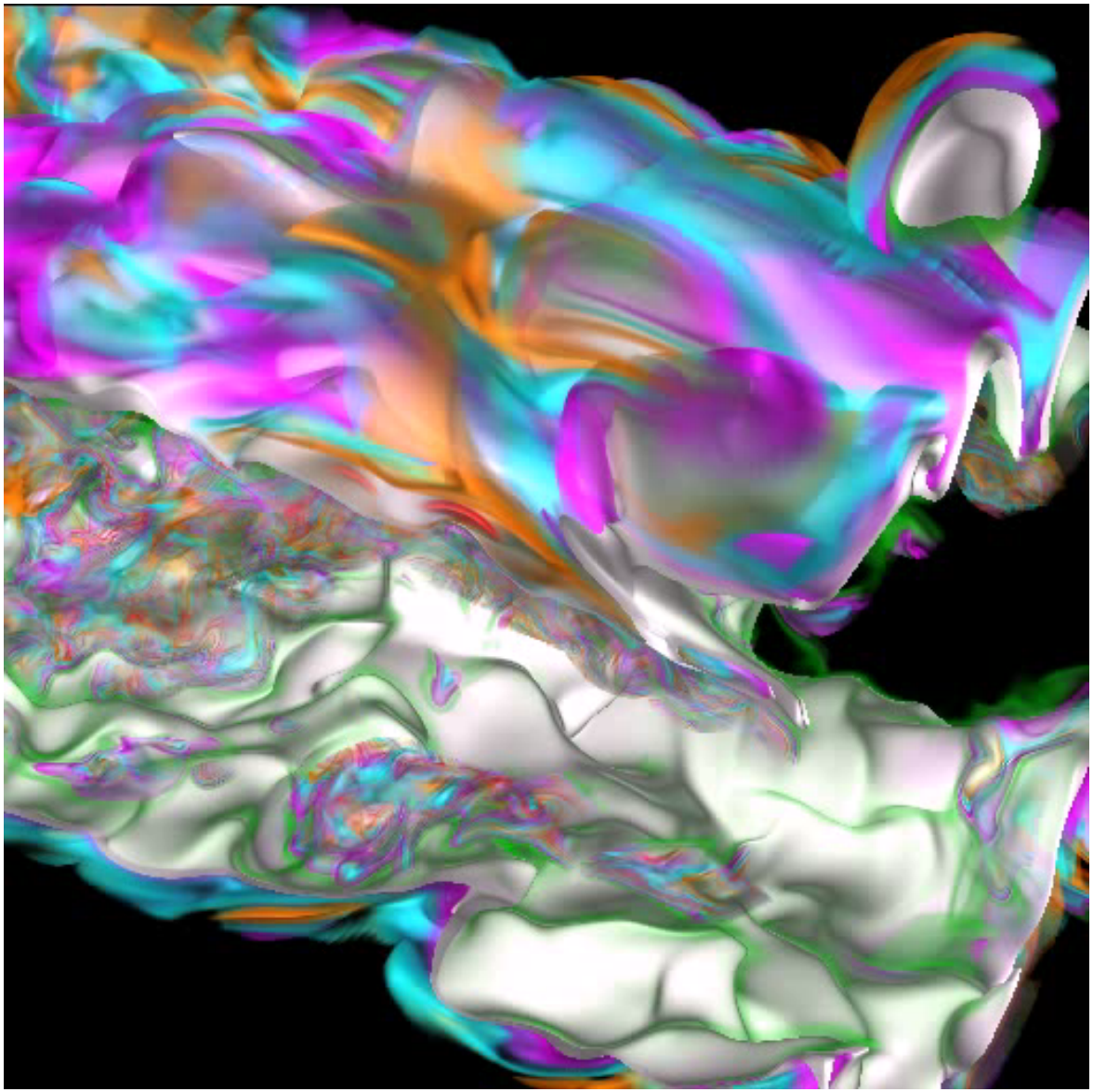
- Design visualization by utilizing scientists' knowledge about the modeled phenomena and their data
- Certain features of interest such as extreme values, physically based properties, geometric object surfaces, etc. may be used to optimize overall visualization
- The results are more efficient & effective visualization
- Knowledge assisted design of all aspects of Vis, from data organization and packing, visual representations, interface and operations, to feedback mechanisms


$HO_2$

# Close-Up View of the Turbulent Flame Surfaces

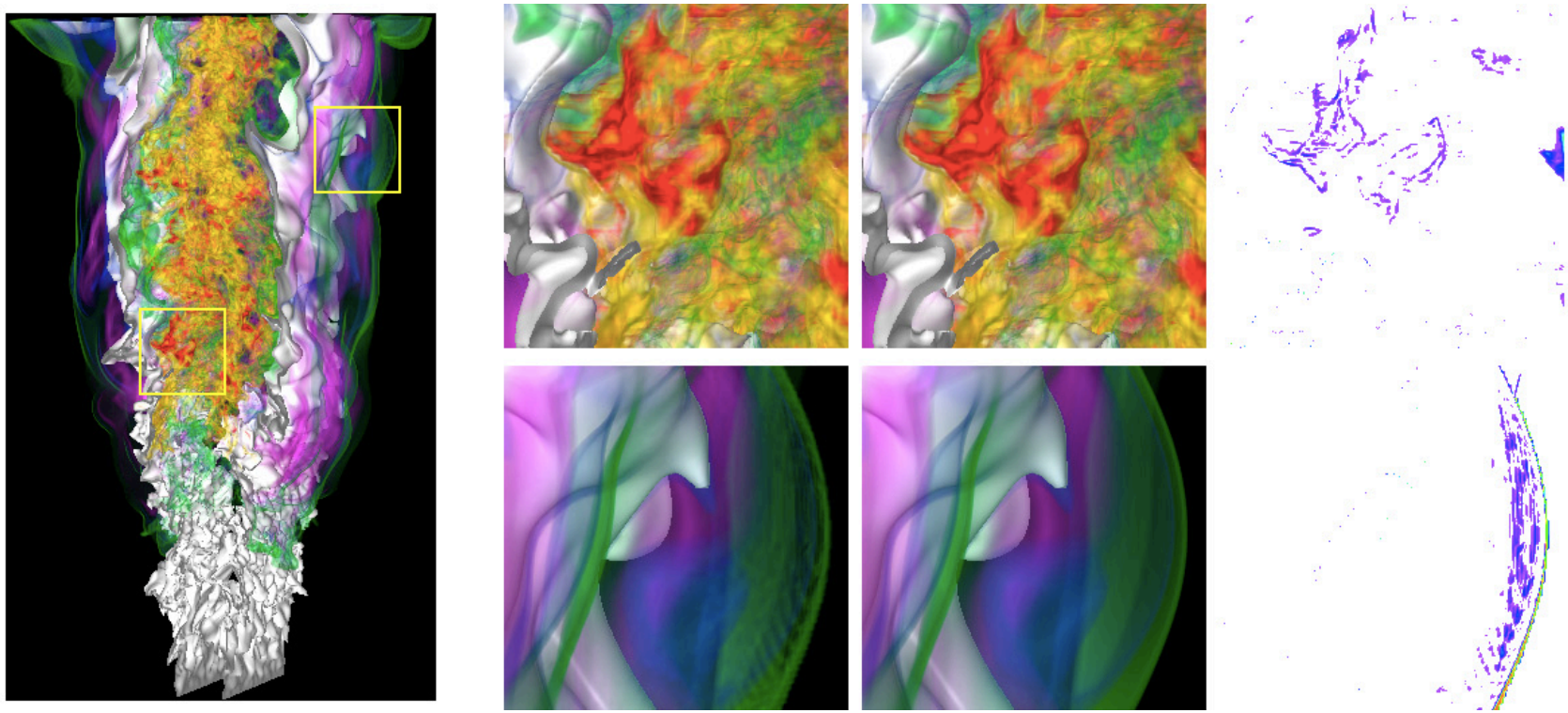






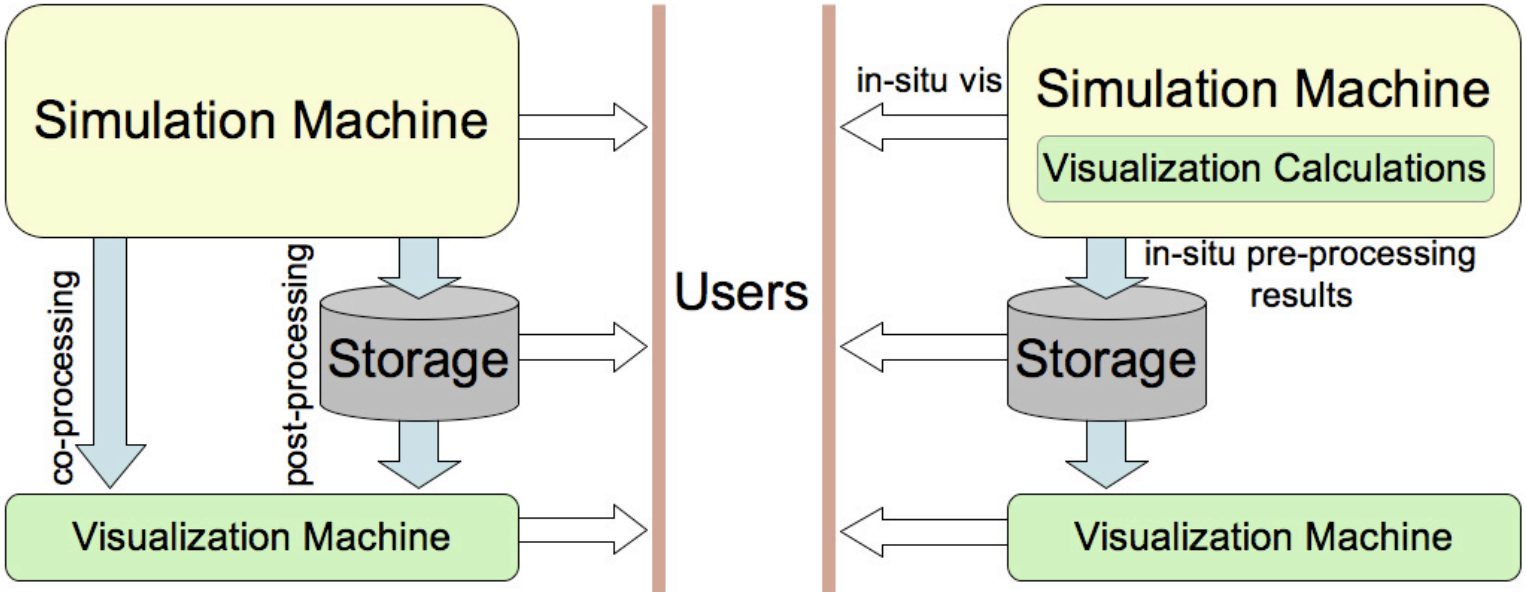
Following the natural coordinate system of a flame, the level-set distance function-based adaptive data reduction algorithms enables us to zoom in and see **for the first time the interaction of small turbulent eddies with the preheat layer of a turbulent flame**, a region that was previously obscured by the multi-scale nature of turbulence.

# Data Reduction Results



# In Situ Visualization

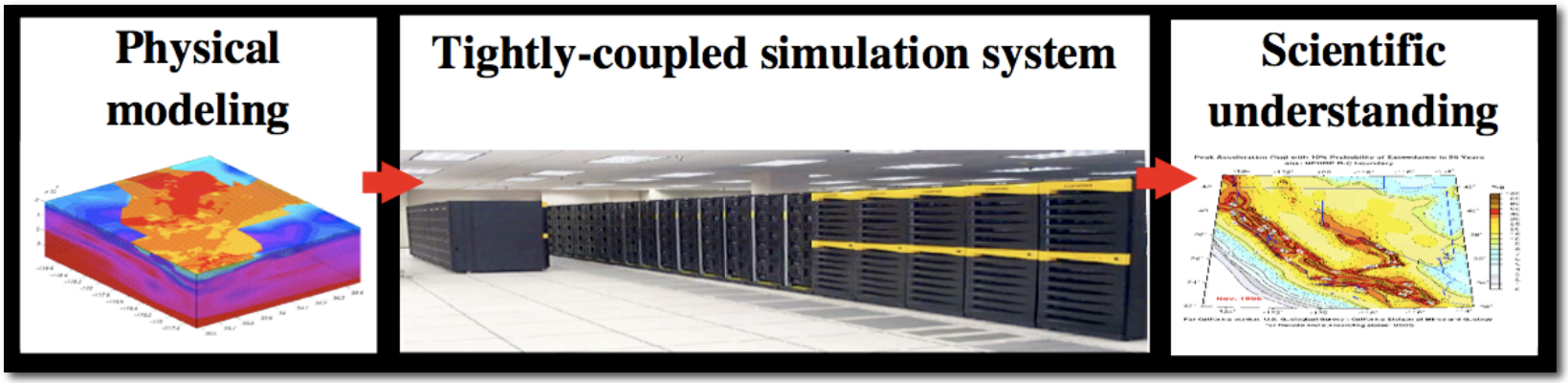
- Reducing and visualizing data in situ as the simulation is running
- The most feasible solution for extreme-scale data analysis



# In Situ Visualization Requirements

- Integration of simulation and visualization codes
- Low memory overhead
- Low computational cost
- Sharing the domain decomposition and data structures
- Scalable parallel visualization algorithms
- Additional requirements for interactive monitoring/steering and different types of visualization

# In Situ Visualization of an Earthquake Simulation




- Run simulation pipelines end-to-end in parallel
- Eliminate scalability bottlenecks
- Execute all components on the same processors
- Simulation-time visualization steering
- Sustained flops increases as the problem size increases
- Winner of SC06 HPC Analytics Challenge on 1024 processors of Cray XT3

# Live Demo at SC06

**Simulation on 1024 processors of PSC Cray XT3**

1994 Northridge Earthquake in Southern California

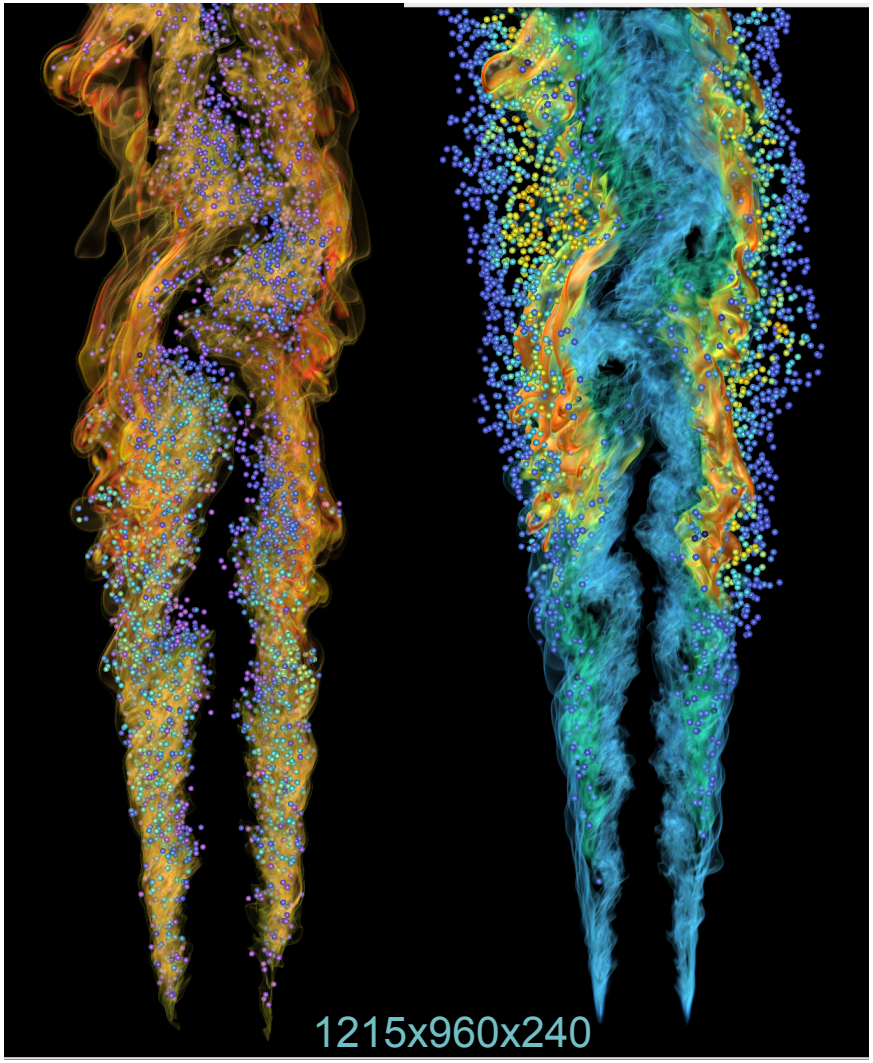
10 Million Elements

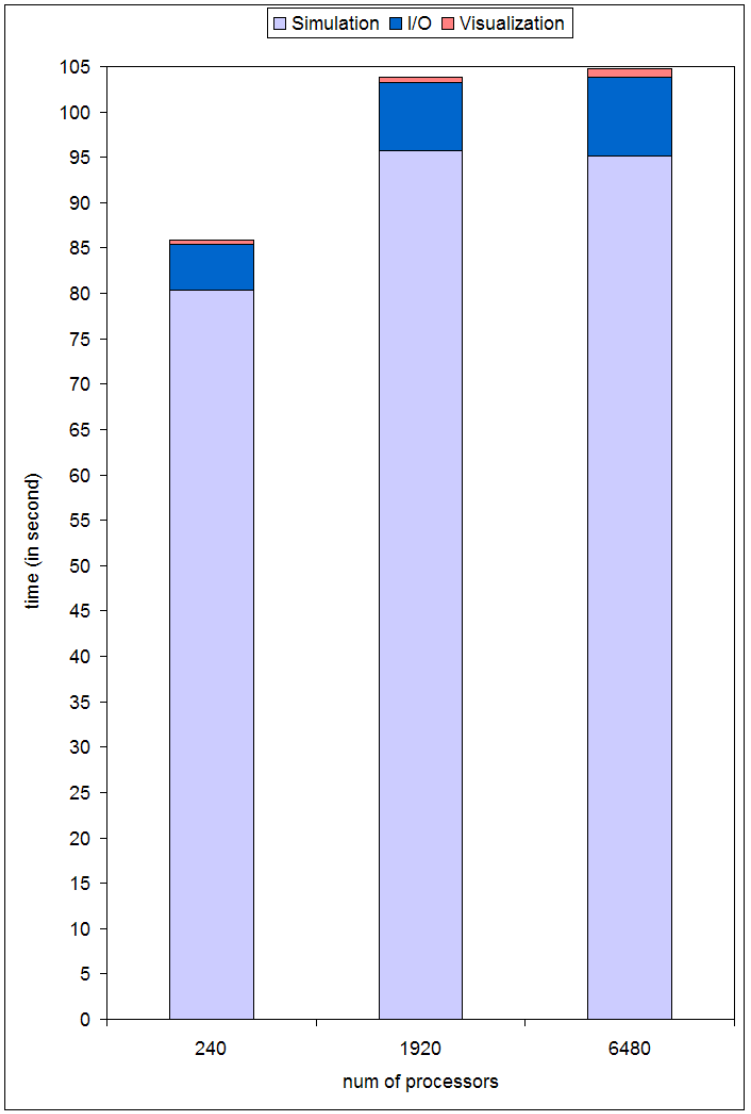
**Steering on a remote laptop computer (1.7 GHz Pentium M, 1 GB memory)**

# In Situ Visualization of a Turbulent Combustion Simulation

- Direct numerical simulation of turbulent combustion (up to 2025x1600x400)
- Visualizing both particle data and volume data
- Using a new highly scalable parallel renderer
- Visualization takes under 1% of overall time
- Using up to 6,480 processors of the Cray XT5 at NCCS/ORNL
- The largest, most scalable in situ visualization ever achieved
- In Situ Vis is feasible!

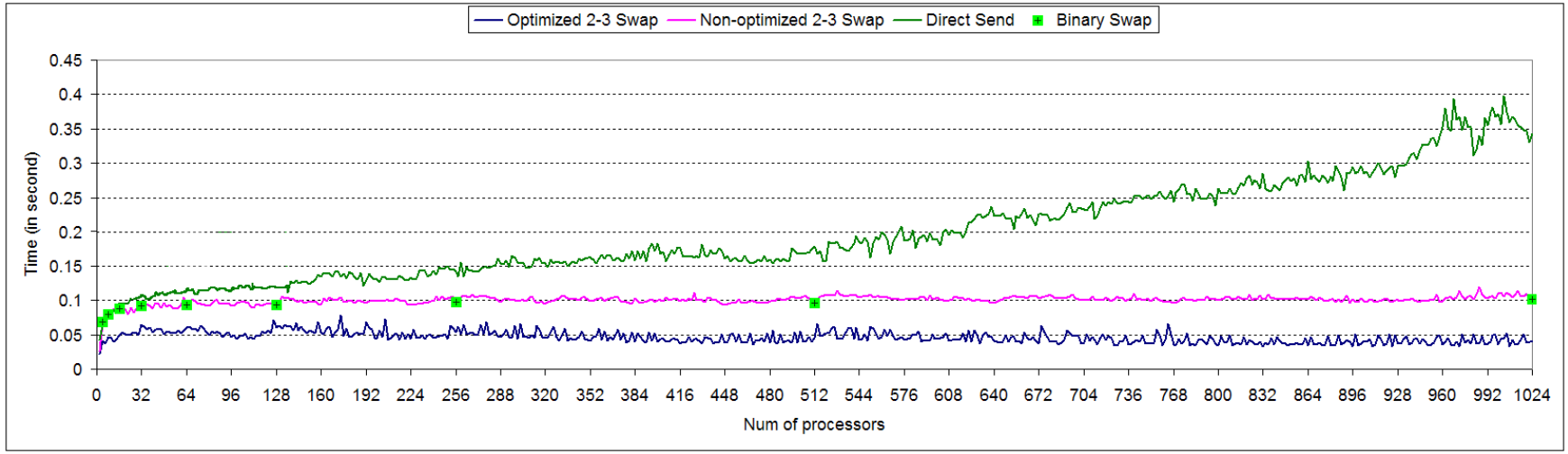
# In Situ Visualization Results



# Parallel Visualization

- Parallel rendering algorithms
- Parallel feature extraction and data reduction
- Multi-GPU computing
- Scalability is the key

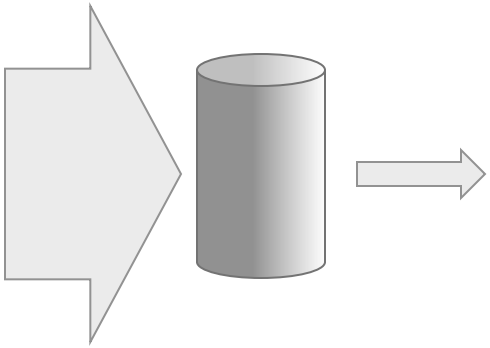


# Remote Visualization


?


Any time
Any where
Any computer

# Remote Visualization

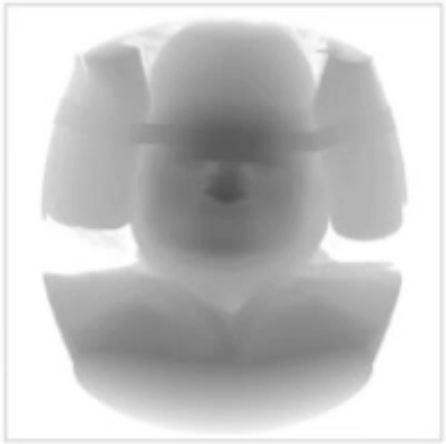




## Explorable Images!

- Spatial domain
- TF space
- Temporal domain
- Rendering space

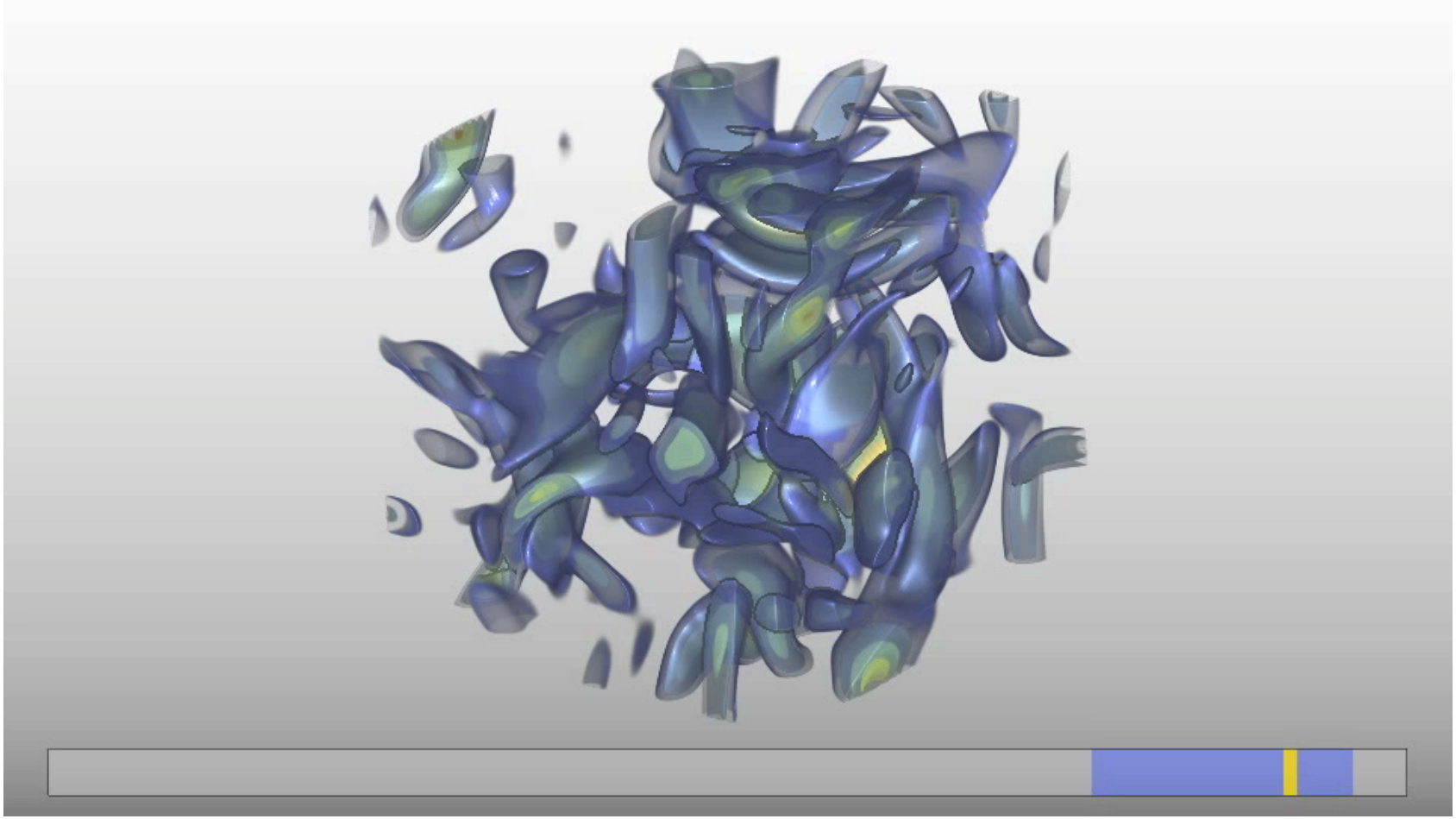
# Explorable Images: Changing View and Shading

Synthesized views from multi-view perspective



CT head

# Explorable Images:
# TF & Temporal Space



# Summary

- Visualization should be an integrated part of the overall scientific discovery process.
- Parallel visualization is absolutely needed for computational science in extreme scale.  Avoid data movement!
- At petascale and exascale, in situ visualization is the most plausible soution.
- Many new advances in visualization will lie in the development of appropriate visual interfaces for data analysis and knowledge discovery.
- Support for remote collaborative and comparative visualization must be developed.

# Kwan-Liu Ma

ma@cs.ucdavis.edu

http://www.cs.ucdavis.edu/~ma


Visualization & Interface Design Innovation